**Decision Tree:**

The machine learning behind this method is to figure out which variable and which threshold to use at every split. One advantage of tree-based methods is that they have no assumptions about the structure of the data and are able to pick up non-linear effects if given sufficient tree depth. We can fit decision trees using our cleaned-up data that has 30 predictor variables and 15884 samples. Model results with Complex Parameter (CP) as 0.0008954804, Root Mean Square Error (RMSE) as 0.48, Rsquare as 0.069 and Mean Absolute Error(MAE) as 0.463.

We tried the model without the “DRGClassification” predictor variable and the accuracy has not changed much. CP selected without DRG is 0.00115872. Accuracy of the training data using decision tree is 60.7%. on testing the model with test data, accuracy was around 59.2%. This model only had a little variation on the RMSE, MAE and Rsquare.

The Variables of importance in creating the tree are number\_inpatient, service\_utilization, number\_emergency, number\_outpatient and DRGClassification.

Decision Tree model did not have much effect based on the gini index criterion with and without the “DRGclassification” variable.